



# Rank related properties, for Basis Pursuit and total variation regularization

François Malgouyres

## ► To cite this version:

François Malgouyres. Rank related properties, for Basis Pursuit and total variation regularization. Signal Processing, 2007, 87 (11), pp.2695-2707. hal-00020801

**HAL Id: hal-00020801**

**<https://hal.science/hal-00020801>**

Submitted on 15 Mar 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Rank related properties, for Basis Pursuit and total variation regularization

F. Malgouyres\*

March 15, 2006

## Abstract

This paper focuses on optimization problems containing an  $l^1$  kind of regularity criterion and a smooth data fidelity term. A general theorem is adapted to this context. It gives an estimate of the distribution law of the "rank" of the solution to the optimization problems, when the initial datum follows a uniform (in a convex compact set) distribution law. It says that asymptotically, solution with a large rank are more and more likely.

The main goal of this paper is to understand the meaning of this notion of rank for some energies which are commonly used in image processing. We study in details the energy whose level sets are defined as the convex hull of finite subset of  $\mathbb{R}^N$  (think Basis Pursuit) and the total variation. For these energies, the notion of rank respectively relates to sparse representation and staircasing.

---

\*LAGA/L2TI, Université Paris 13, [malgouy@math.univ-paris13.fr](mailto:malgouy@math.univ-paris13.fr)

# 1 Introduction

This paper focuses on some properties of energies whose level sets are polyhedral. To be specific, we consider an energy of the form

$$E(u) = \sup_{\Psi \in \mathcal{D}} \langle u, \Psi \rangle, \quad (1)$$

where  $u \in \mathbb{R}^N$ ,  $\mathcal{D}$  is dictionary in  $\mathbb{R}^N$  and  $\langle \cdot, \cdot \rangle$  stands for the scalar product in  $\mathbb{R}^N$ .

The interest for such energies comes from their wide use in today scientific calculus. Almost all our examples come from image processing alone. In that field, such energies are used in : total variation regularization (for references, see the introduction of Section 4), in Basis Pursuit and alike algorithms (see references in Section 3), wavelet soft thresholding (see [1, 2], where soft-thresholding is proved to be the solution to optimization problems involving energies whose level sets are polyhedral), image compression (see [3], where JPEG and JPEG2000 are interpreted in terms of solutions to optimization problems involving an energy whose level sets are polyhedral),...

An analysis of such optimization problems is presented in [4]. In few words, it says that asymptotically, as  $E(u)$  becomes small ( $u$  is the solution to the optimization problem),  $u$  is more and more likely to have a large rank (the precise definition of the rank is given in Section 2). However, the meaning of the notion of rank might not be obvious for some energies listed above.

The purpose of this paper is to establish this meaning for two different kinds of energies : those whose levels sets is easily expressed as the convex hull of a finite subset of  $\mathbb{R}^N$  and the total variation. In simple words, we study the meaning of the notion of rank for energies of the  $l^1$  kind.

In fact, the notion of rank is quite explicit when the energy is already expressed under the form (1). Its importance has already been highlighted in [3], where it is shown that the compression standards JPEG and JPEG2000 can be considered particular cases of more general compression scheme for which the number of coefficients that need to be coded is the rank of the solution to an optimization problem involving an energy whose level sets are polyhedra.

The paper sketches as Follows :

- In Section 2, we summarize the notions and results established in [4].
- In Section 3, we study energies whose level sets are defined as the convex hull of a finite subset  $\mathcal{B} \subset \mathbb{R}^N$ . Examples of use of such energies are found in Basis Pursuit algorithms. We show that, for any  $u \in \mathbb{R}^N$  and such an energy,  $u$  can be expressed with  $N - rk(u) + 1$  non-zero coordinates, where  $rk(u)$  denotes the rank of  $u$ . In this context, the results established in [4] express that, in many situations, algorithms involving such an regularity criterion are very likely to provide a solution which is represented with only few coordinates in  $\mathcal{B}$ .

- In Section 4, we study an approximation of the total variation whose level sets are polyhedral. We establish some links between the rank of an element  $u \in \mathbb{R}^{N^2}$  and the size of the set

$$I_2 = \{(i, j) \in \mathbb{R}^{N^2}, \nabla u_{i,j} = (0, 0)\}.$$

These results suggest the staircasing (which is usually quantified by the size of  $I_2$ ) relates to the notion of rank. However, the notion of rank, by itself, does not permit to quantify the staircasing artifact. A deeper analysis is probably needed. This shows the limitations of the notion of rank for energies like the total variation.

The paper is organized so that a reader can skip Section 3 or 4. They are completely independent and both contain a short introduction to the concerned research area.

## 2 A rank likelihood estimate

As a particular case, [4] studies an optimization problem of the form

$$(P) : \begin{cases} \text{Minimize } D(w - u_0) \\ \text{under the constraint } E(w) \leq \tau \end{cases}$$

for a datum  $u_0 \in \mathbb{R}^N$ ,  $E(u_0) > \tau > 0$  and two functions  $E$  and  $D$ , defined over  $\mathbb{R}^N$ . In the text, we will refer to  $D$  as the "data fidelity term" and to  $E$  as the "regularization term".

In [4], it is assumed that  $D$  is a regular (i.e. continuously differentiable away from 0) norm. (The boundary of the level sets of  $D$  also needs to be slightly "curved", we recommend to see [4] for details.) The function  $E$  is assumed to be a norm whose level-sets are polyhedra.

Notice that in general such optimization models are written under the form

$$\begin{cases} \text{Minimize } E(w) \\ \text{under the constraint } D(w - u_0) \leq \tau' \end{cases}$$

for  $D(-u_0) > \tau' > 0$ , or

$$\text{Minimize } E(w) + \lambda D(w - u_0),$$

for  $\lambda > 0$ .

However, under the hypotheses of the theorem below, those optimization problems are equivalent to  $(P)$ . Their solutions are indeed always characterized by

$$-\lambda' D(w - u_0) \in \partial E(w),$$

for an adequate  $\lambda' > 0$ . ( $\partial E$  denotes the subgradient of  $E$ .)

Coming back to  $(P)$ , we know (see [4]) that, because of the hypotheses on  $E$ , there exists a finite subset  $(\varphi_i)_{i \in I} \subset \mathbb{R}^N$  such that, for all  $u \in \mathbb{R}^N$ ,

$$E(u) = \sup_{i \in I} \langle u, \varphi_i \rangle. \quad (2)$$

Defining, for all  $u \in \mathbb{R}^N$ ,

$$J(u) \triangleq \{i \in I, E(u) = \langle u, \varphi_i \rangle\},$$

we can prove (see [4]) that

$$rk(u) \triangleq \dim(\text{Span}((\varphi_j)_{j \in J(u)}))$$

does not depend on the choice of  $(\varphi_i)_{i \in I}$ , in (2). We call  $rk(u)$ , the rank of  $u$ .

We denote, for a function  $f$  and  $\tau > 0$ , the level set of a function  $f$  by

$$\mathcal{L}_f(\tau) = \{u, f(u) \leq \tau\}.$$

**Theorem 1** *Let  $f$  be a norm over  $\mathbb{R}^N$  and  $U$  be a random variable whose distribution law is uniform in a set  $\mathcal{A}$  satisfying*

$$\mathcal{L}_f(\tau_1) \subset \mathcal{A} \subset \mathcal{L}_f(\tau_2),$$

for  $\tau_2 \geq \tau_1 > 0$ .

*There exists non-negative real numbers  $(C_K)_{1 \leq K \leq N}$  such that, for all  $K \in \{1, \dots, N\}$ ,*

$$C_K \tau^{N-K} \frac{\tau_1^K}{\mathbb{L}(\mathcal{A})} + o(\tau^{N-K}) \leq \mathbb{P}(rk(\overline{U}) = K) \leq C_K \tau^{N-K} \frac{\tau_2^K}{\mathbb{L}(\mathcal{A})} + o(\tau^{N-K})$$

as  $\tau$  goes to 0, where  $\mathbb{L}(\mathcal{A})$  is the Lebesgue measure of  $\mathcal{A}$  and  $\overline{U}$  denotes the solution to (P) when  $u_0$  is a realization of  $U$ . (The hypotheses on  $E$  and  $D$  are given right after the definition of (P).)

This theorem quantifies the probability for the solution to (P) to have a given rank, when  $\tau$  is small. It tells us that, in this case,  $rk(u)$  is very likely to be large. This holds under quite general assumptions on the data distribution law and for a large class of data fidelity term. Notice that, usual image processing models, the goal of models like (P) is to provide a solution whose rank is small (see [3], and the next sections).

However, for a fixed regularization term, we can increase the likelihood of getting a small rank through the improvement of the constants  $C_K$ . When expressing  $C_K$  (see [4]), we see that they strongly depend on  $D$ . Doing so, we can expect improvements but cannot expect to go beyond the estimation given in Theorem 1. The only way to go beyond those estimations would be to chose  $D$  that fails to comply with its hypotheses.

### 3 The Basis Pursuit regularization term

We consider in this section the energy defined for  $u \in \mathbb{R}^N$  by

$$E(u) = \min_{(\alpha_j)_{j \in J} \in \mathcal{C}(u)} \sum_{j \in J} \alpha_j \quad (3)$$

where

$$\mathcal{C}(u) = \{(\alpha_j)_{j \in J}, \text{ such that } \forall j \in J, \alpha_j \geq 0 \text{ and } u = \sum_{j \in J} \alpha_j \varphi_j\}$$

where  $\mathcal{B} = (\varphi_j)_{j \in J}$  is a finite subset of  $\mathbb{R}^N$  such that

1. for all  $j \in J$ ,  $\|\varphi_j\|_2 = 1$ ,
2. for all  $(i, j) \in J^2$ , if  $i \neq j$ ,  $\varphi_i \neq \varphi_j$ ,
3. and for all  $u \in \mathbb{R}^N$ ,  $\mathcal{C}(u) \neq \emptyset$ .

We will often abuse of the notation  $(\alpha_i)_{i \in I} \in \mathcal{C}(u)$ , for  $I \subset J$ . In such a case, one should understand that the numbers  $(\alpha_j)_{j \in J \setminus I}$  are implicitly set to 0.

Notice that we could easily get rid of the last condition on  $\mathcal{B}$ . In this case  $E$  is only defined over

$$\left\{ u \in \mathbb{R}^N, \exists (\alpha_j)_{j \in J} \in \mathbb{R}^J \text{ such that } u = \sum_{j \in J} \alpha_j \varphi_j \text{ and } \forall j \in J, \alpha_j \geq 0 \right\}.$$

For simplicity, we will always consider the condition 3 holds though.

One motivation for considering such an energy is that it is already used in Basis Pursuit [5], for source separation [6], and for feature selection in classification [7]. Let us also mention the theoretical work in [8, 9] which specifies the geometry of  $\mathcal{B}$  so that the  $(\alpha_j)_{j \in J}$  solving  $E(u)$  is the sparsest in  $\mathcal{C}(u)$ . The Basis pursuit norm (the one considered in those paper) is of the form  $E$  for a dictionary  $\mathcal{B} = \tilde{\mathcal{B}} \cup \{-\varphi, \varphi \in \tilde{\mathcal{B}}\}$ , where  $\tilde{\mathcal{B}}$  the dictionary considered in those papers.

Another important motivation for studying energies of the form  $E$  is that its level sets are scaled version of the convex hull of  $\mathcal{B}$ . Indeed, under the hypotheses above,  $E$  is clearly a norm and we have

$$\{u, E(u) \leq 1\} = \left\{ u = \sum_{j \in J} \alpha_j \varphi_j, \text{ where } \forall j \in J, \alpha_j \geq 0 \text{ and } \sum_{j \in J} \alpha_j \leq 1 \right\},$$

so

$$\mathcal{L}_E(1) = \left\{ u = \sum_{j \in J} \alpha_j \varphi_j + \alpha_0 0, \text{ where } \alpha_0 \geq 0, \forall j \in J, \alpha_j \geq 0 \right. \\ \left. \text{and } \alpha_0 + \sum_{j \in J} \alpha_j = 1 \right\}$$

As a consequence,  $\mathcal{L}_u(1)$  is the convex hull of  $\mathcal{B} \cup \{0\}$  (see [10], Cor. 2.3.1, pp. 12). But, because of the condition 3, 0 belongs to the interior of  $\mathcal{L}_u(1)$  and therefore the convex hull of  $\mathcal{B} \cup \{0\}$  equals the convex hull of  $\mathcal{B}$ . Of course, since  $E$  is a norm, for any  $\tau > 0$ ,  $\mathcal{L}_E(\tau)$  is a scaled version of the convex hull of  $\mathcal{B}$ .

As a consequence, the level sets of  $E$  are polyhedral (see [10], Th. 19.1, pp. 171) and the results described in Section 2 can be applied to such an energy.

Notice that the representation of convex polyhedron as the convex hull of a finite set of elements is classical (see [10], Section 18, pp. 162). This, by itself, is an important motivation for studying energies whose level sets are polyhedron in this framework.

The question we would like to address here is the relation between the rank of an element  $u \in \mathbb{R}^N$  and some properties on  $u$ . Heuristically, it seems clear that as the rank of  $u$  grows, we need less and less  $\varphi_j$  to represent  $u$ .

Let us first define the set  $\mathcal{D}$ , such that

$$E(u) = \sup_{\Psi \in \mathcal{D}} \langle u, \Psi \rangle.$$

Here and all along the section,  $\langle \cdot, \cdot \rangle$  stands for the scalar product in  $\mathbb{R}^N$ .

**Proposition 1** *Let  $u \in \mathbb{R}^N$  and let  $(\alpha'_j)_{j \in J} \in \mathcal{C}(u)$  be such that*

$$E(u) = \sum_{j \in J} \alpha'_j.$$

*We denote*

$$I' = \{j \in J, \alpha'_j \neq 0\}$$

*It is possible to extract, from  $I'$ , a set  $I$  such that  $(\varphi_i)_{i \in I}$  is an independent system of vectors and there exists  $(\alpha_i)_{i \in I} \in \mathcal{C}(u)$  such that for all  $i \in I$ ,  $\alpha_i > 0$ , and*

$$E(u) = \sum_{i \in I} \alpha_i.$$

*Proof.* First, if  $I'$  is such that  $(\varphi_i)_{i \in I'}$  is not an independent system of vectors, there exists  $i_0 \in I'$  such that

$$\varphi_{i_0} = \sum_{i \in I' \setminus \{i_0\}} \alpha_i^{i_0} \varphi_i$$

So, we can deduce from the hypotheses of the proposition that, for all  $\lambda \in \mathbb{R}$ ,

$$\begin{aligned} u &= \sum_{i \in I'} \alpha'_i \varphi_i \\ &= \sum_{i \in I'} \alpha'_i \varphi_i + \lambda \left( \varphi_{i_0} - \sum_{i \in I' \setminus \{i_0\}} \alpha_i^{i_0} \varphi_i \right) \\ &= \sum_{i \in I' \setminus \{i_0\}} (\alpha'_i - \lambda \alpha_i^{i_0}) \varphi_i + (\alpha'_{i_0} + \lambda) \varphi_{i_0}. \end{aligned}$$

We denote, for  $j \in J$

$$\alpha_j^\lambda = \begin{cases} \alpha'_j - \lambda \alpha_j^{i_0} & , \text{ if } j \in I' \setminus \{i_0\} \\ \alpha_{i_0} + \lambda & , \text{ if } j = i_0 \\ 0 & , \text{ otherwise.} \end{cases}$$

Notice that, for  $\lambda$  small enough ( $\lambda$  can be positive or negative),  $(\alpha_j^\lambda)_{j \in J} \in \mathcal{C}(u)$ , and

$$\sum_{j \in J} \alpha_j^\lambda = \sum_{i \in I'} \alpha'_i + \lambda \left( 1 - \sum_{i \in I' \setminus \{i_0\}} \alpha_i^{i_0} \right).$$

So, since  $(\alpha_j)_{j \in J}$  is solution to (3), we necessarily have

$$\sum_{i \in I' \setminus \{i_0\}} \alpha_i^{i_0} = 1.$$

As a consequence, as long as, for all  $j \in J$ ,  $\alpha_j^\lambda \geq 0$ ,  $(\alpha_j^\lambda)_{j \in J}$  is also solution to (3). We take  $\lambda_0$  equal to the largest negative value such that, for all  $j \in J$ ,  $\alpha_j^\lambda \geq 0$  (it is not difficult to see it exists and that  $\lambda_0 < 0$ ) and we obtain  $(\alpha_j^{\lambda_0})_{j \in J} \in \mathcal{C}(u)$  such that

$$E(u) = \sum_{i \in I'} \alpha_i^{\lambda_0}$$

and

$$\#\{j \in J, \alpha_j^{\lambda_0} \neq 0\} < \#I'.$$

We can repeat the same procedure as long as the so obtained set of indexes does not define an independent system of vectors. We denote the result  $I$ . Obviously,  $I$  satisfies the conditions of Proposition 1.  $\square$

**Definition 1** We call face any subset  $I \subset J$  such that  $(\varphi_i)_{i \in I}$  is a basis and there exists  $u \in \mathbb{R}^N$  which cannot be expressed as a linear combination of strictly less than  $N$  elements of  $\mathcal{B}$ , such that there exists  $(\alpha_i)_{i \in I} \in \mathcal{C}(u)$  satisfying

$$E(u) = \sum_{i \in I} \alpha_i.$$

We denote  $\mathcal{F}$  the set of all the faces.

For any  $I \in \mathcal{F}$ , we denote  $\Psi_I$  the only solution to

$$\forall i \in I, \langle \Psi_I, \varphi_i \rangle = 1.$$

Notice that for any  $u \in \mathbb{R}^N$ , that cannot be expressed as a linear combination of strictly less than  $N$  elements of  $\mathcal{B}$ , Proposition 1 guaranties there exists a face  $I \in \mathcal{F}$  such that

$$E(u) = \sum_{i \in I} \alpha_i,$$



for a given  $(\alpha_i)_{i \in I} \in \mathcal{C}(u)$ . (It is, by the way, clear that for all  $i \in I$ ,  $\alpha_i > 0$ .)

Notice also that

$$E(u) = \sum_{i \in I} \alpha_i = \sum_{i \in I} \alpha_i \langle \varphi_i, \Psi_I \rangle = \langle u, \Psi_I \rangle.$$

In fact, we can prove a little more.

**Proposition 2** *Let  $I$  be a face. For any*

$$v \in \left\{ \sum_{i \in I} \alpha_i \varphi_i, \forall i \in I, \alpha_i > 0 \right\},$$

*there exists  $(\alpha_i)_{i \in I} \in \mathcal{C}(v)$  and*

$$E(v) = \sum_{i \in I} \alpha_i = \langle v, \psi_I \rangle.$$

*Proof.* First, since  $I$  is a face, there exists  $u \in \mathbb{R}^N$  such that,  $u$  cannot be expressed as a linear combination of strictly less than  $N$  elements of  $\mathcal{B}$  and there exists  $(\alpha_i)_{i \in I} \in \mathcal{C}(u)$  and

$$E(u) = \sum_{i \in I} \alpha_i = \langle u, \psi_I \rangle.$$

Notice that, for any  $i \in I$ ,  $\alpha_i > 0$ .

Let us observe that, since  $I$  is a face, for any  $j \in J \setminus I$ , there exists  $(\alpha_i^j)_{i \in I} \in \mathbb{R}^I$  such that

$$\varphi_j = \sum_{i \in I} \alpha_i^j \varphi_i.$$

So, for any  $\lambda \geq 0$ ,

$$\begin{aligned} u &= \sum_{i \in I} \alpha_i \varphi_i + \lambda \left( \varphi_j - \sum_{i \in I} \alpha_i^j \varphi_i \right) \\ &= \sum_{i \in I} (\alpha_i - \lambda \alpha_i^j) \varphi_i + \lambda \varphi_j. \end{aligned}$$

Moreover, since for all  $i \in I$ ,  $\alpha_i > 0$ , for  $\lambda > 0$  and small enough, we have

$$(\alpha'_i)_{i \in J} \in \mathcal{C}(u),$$

where

$$\alpha'_i = \begin{cases} \alpha_i - \lambda \alpha_i^j & , \text{ if } i \in I \\ \lambda & , \text{ if } i = j \\ 0 & , \text{ otherwise.} \end{cases}$$

As a consequence,

$$\begin{aligned}\sum_{i \in I} \alpha_i &\leq \sum_{j \in J} \alpha'_j \\ &\leq \sum_{i \in I} \alpha_i + \lambda \left( 1 - \sum_{i \in I} \alpha_i^j \right).\end{aligned}$$

Finally, for any  $j \in J \setminus I$ ,

$$\sum_{i \in I} \alpha_i^j \leq 1. \quad (4)$$

Let us consider  $(\alpha_i^v)_{i \in I}$ , such that for all  $i \in I$ ,  $\alpha_i^v > 0$ , and

$$v = \sum_{i \in I} \alpha_i^v \varphi_i.$$

For any  $(\beta_j)_{j \in J} \in \mathcal{C}(v)$ , it is not difficult to see that, for any  $i \in I$ ,

$$\alpha_i^v = \sum_{j \in J} \beta_j \alpha_i^j.$$

(Remember indeed, that, since  $I$  is a face,  $(\varphi_i)_{i \in I}$  is a basis.)

So, since (4) holds,

$$\begin{aligned}\sum_{i \in I} \alpha_i^v &= \sum_{j \in J} \beta_j \sum_{i \in I} \alpha_i^j, \\ &\leq \sum_{j \in J} \beta_j.\end{aligned}$$

Therefore

$$E(v) = \sum_{i \in I} \alpha_i^v = \langle v, \psi_I \rangle.$$

□

**Theorem 2** For any  $u \in \mathbb{R}^N$ ,

$$E(u) = \sup_{I \in \mathcal{F}} \langle u, \Psi_I \rangle.$$

*Proof.* Let us first consider  $u \in \mathbb{R}^N$  such that  $u$  cannot be expressed as a linear combination of strictly less than  $N$  elements of  $\mathcal{B}$ .

From Proposition 1 and Definition 1, we know there exists  $I \in \mathcal{F}$  such that

$$E(u) = \langle u, \Psi_I \rangle.$$

So, we know that

$$E(u) \leq \sup_{I' \in \mathcal{F}} \langle u, \Psi_{I'} \rangle.$$

In order to prove the converse statement, let us consider a face  $I' \in \mathcal{F}$ . Since it is a face, we know there exists  $v \in \mathbb{R}^N$  such that

$$E(v) = \langle v, \Psi_{I'} \rangle$$

and  $v$  cannot be expressed as a linear combination of strictly less than  $N$  elements of  $\mathcal{B}$ .

Let us consider, for  $\lambda \in [0, 1]$ ,

$$u^\lambda = \lambda u + (1 - \lambda)v.$$

Proposition 2 guaranties that for  $\lambda$  small enough

$$E(u^\lambda) = \langle u^\lambda, \Psi_{I'} \rangle.$$

So, for such a  $\lambda$ ,

$$\begin{aligned} \lambda \langle u, \Psi_{I'} \rangle + (1 - \lambda) \langle v, \Psi_{I'} \rangle &= \langle u^\lambda, \Psi_{I'} \rangle \\ &= E(\lambda u + (1 - \lambda)v) \\ &\leq \lambda E(u) + (1 - \lambda)E(v) \\ &\leq \lambda \langle u, \Psi_I \rangle + (1 - \lambda) \langle v, \Psi_{I'} \rangle. \end{aligned}$$

This guaranties that

$$\langle u, \Psi_{I'} \rangle \leq \langle u, \Psi_I \rangle.$$

This guaranties that for any  $u \in \mathbb{R}^N$  such that  $u$  cannot be expressed as a linear combination of strictly less than  $N$  elements of  $\mathcal{B}$

$$E(u) = \sup_{I \in \mathcal{F}} \langle u, \Psi_I \rangle.$$

The final result follows from the remark

$$\{u \in \mathbb{R}^N, u \text{ cannot be expressed as a linear combination of strictly less than } N \text{ elements of } \mathcal{B}\}$$

is dense in  $\mathbb{R}^N$  and the fact that both  $E$  and  $\sup_{I \in \mathcal{F}} \langle \cdot, \Psi_I \rangle$  are continuous.  $\square$

Notice that, one of the consequences of Theorem 2 is that for all  $j \in J$  and all  $I \in \mathcal{F}$ ,

$$\langle \varphi_j, \Psi_I \rangle \leq 1.$$

We can now rewrite, for  $u \in \mathbb{R}^N$ , the definition of  $rk(u)$  for the energy  $E$ . We have

$$rk(u) = \dim \left( \text{Span}(\Psi_I)_{I \in \mathcal{F}(u)} \right),$$

for

$$\mathcal{F}(u) = \{I \in \mathcal{F}, \langle u, \Psi_I \rangle = E(u)\}.$$

Let us also recall a definition that was given in [4]. We denote

$$\bar{u} = \{v \in \mathbb{R}^N, \forall I \in \mathcal{F}, \langle v, \Psi_I \rangle = E(v) \Leftrightarrow I \in \mathcal{F}(u), \text{ and } E(u) = E(v)\},$$

for  $u \in \mathbb{R}^N$ .

As is already remarked in [4],  $\bar{u}$  is an affine manifold whose dimension is  $N - rk(u)$ .

Let us finally denote

$$I_0 = \{j \in J, \forall I \in \mathcal{F}(u), \langle \varphi_j, \Psi_I \rangle = 1\}.$$

The following proposition holds

**Proposition 3** *For any  $u \in \mathbb{R}^N$ ,*

$$\bar{u} \subset \left\{ \sum_{i \in I_0} \alpha_i \varphi_i, \forall i \in I_0, \alpha_i \geq 0, \sum_{i \in I_0} \alpha_i = E(u) \right\}. \quad (5)$$

and

$$\left\{ \sum_{i \in I_0} \alpha_i \varphi_i, \forall i \in I_0, \alpha_i > 0, \text{ and } \sum_{i \in I_0} \alpha_i = E(u) \right\} \subset \bar{u}$$

*Proof.* In order to prove the First inclusion, we consider  $v \in \bar{u}$ . Proposition 1 guaranties there exists  $I_1 \subset J$  such that  $(\varphi_i)_{i \in I_1}$  is an independent system of vectors and there exists  $(\alpha_i)_{i \in I_1} \in \mathcal{C}(v)$  such that for all  $i \in I_1$ ,  $\alpha_i > 0$  and

$$E(v) = \sum_{i \in I_1} \alpha_i.$$

If there exists  $i_0 \in I_1 \setminus I_0$ , there exists  $I \in \mathcal{F}(u)$  such that

$$\langle \varphi_{i_0}, \Psi_I \rangle < 1.$$

Therefore, since  $\alpha_{i_0} > 0$  (the other  $\alpha_i$  are positive),

$$\begin{aligned} \langle v, \Psi_I \rangle &= \sum_{i \in I_1} \alpha_i \langle \varphi_i, \Psi_I \rangle \\ &< \sum_{i \in I_1} \alpha_i = E(v) \end{aligned}$$

This contradicts  $v \in \bar{u}$ . So  $I_1 \subset I_0$ , and the second inclusion is proved.

In order to prove the second inclusion, let  $(\alpha_i)_{i \in I_0} \in \mathbb{R}^{I_0}$  be such that for all  $i \in I_0$ ,  $\alpha_i > 0$  and

$$\sum_{i \in I_0} \alpha_i = E(u).$$

We denote

$$v = \sum_{i \in I_0} \alpha_i \varphi_i. \quad (6)$$

For any  $I \in \mathcal{F}(u)$ ,

$$\begin{aligned} \langle v, \Psi_I \rangle &= \sum_{i \in I_0} \alpha_i \langle \varphi_i, \Psi_I \rangle \\ &= E(u). \end{aligned}$$

Let us consider  $I \notin \mathcal{F}(u)$ . We know that

$$\langle u, \Psi_I \rangle < E(u). \quad (7)$$

Now, since we already proved (5) and  $u$  obviously belong to  $\overline{u}$ , there exists  $(\alpha_i)_{i \in I_0}$ , such that for all  $i \in I_0$ ,  $\alpha_i \geq 0$ ,

$$u = \sum_{i \in I_0} \alpha_i \varphi_i,$$

and

$$E(u) = \sum_{i \in I_0} \alpha_i.$$

So, (7) becomes

$$\sum_{i \in I_0} \alpha_i \langle \varphi_i, \Psi_I \rangle < \sum_{i \in I_0} \alpha_i,$$

which guaranties there exists  $i_0 \in I_0$  such that

$$\langle \varphi_{i_0}, \Psi_I \rangle < 1.$$

Finally, we know that, for any  $I \notin \mathcal{F}(u)$ , there exists  $i_0 \in I_0$  such that

$$\langle \varphi_{i_0}, \Psi_I \rangle < 1.$$

Since, in (6),  $\alpha_{i_0} > 0$  (the other  $\alpha_i$  are positive),

$$\begin{aligned} \langle v, \Psi_I \rangle &= \sum_{i \in I_0} \alpha_i \langle \varphi_i, \Psi_I \rangle \\ &< E(u). \end{aligned}$$

So  $v \in \overline{u}$ .

□

As a consequence, we have for any  $u \in \mathbb{R}^N$ ,

$$\begin{aligned} N - rk(u) &= \dim(\overline{u}) \\ &= \dim(\text{Span}(\varphi_i)_{i \in I_0}) - 1 \end{aligned} \quad (8)$$

**Corollary 1** *For any  $u \in \mathbb{R}^N$ , there exist  $I \subset J$  and  $(\alpha_i)_{i \in I} \in \mathcal{C}(u)$  such that*

$$\#I \leq N - rk(u) + 1.$$

*Proof.* This is a direct consequence of Proposition 3, Proposition 1 and (8).  $\square$

Proposition 3 and the above Corollary tell us that the Basis Pursuit algorithm make the link between the existence of a sparse representation for a given  $u \in \mathbb{R}^N$ , in  $\mathcal{B}$ , and the rank of  $u$ . If the rank is large, there exists a sparse representation of  $u$ , in  $\mathcal{B}$ . Moreover, Proposition 1 tells us that, modulo the application of a simple algorithm (see the proof of Proposition 1, where the algorithm is explicitly given), basis pursuit provides us with a description of  $u$  with  $N - rk(u) + 1$  coordinates.

In this framework, Theorem 1 tells us that as  $\tau$  goes to 0, the solution to  $(P)$  is more and more likely to be sparse.

Finally, notice that the question of finding the sparsest representation for  $u$ , in  $\mathcal{B}$ , is not solved in general by a basis pursuit algorithm (see [8, 9]). Notice however, that, for a fixed  $\mathcal{B}$ , the set of elements  $u$  that can be expressed with strictly less than  $N$  independent elements of  $\mathcal{B}$  is obviously of Lebesgue measure 0.

An interesting observation is that, similarly, the set of all the elements in  $\overline{u}$  which can be expressed with strictly less than  $N - rk(u) + 1$  independent elements of  $\mathcal{B}$  is of  $N - rk(u)$  Hausdorff measure 0. Therefore, when solving  $(P)$ , for an initial datum as in Theorem 1, we can expect the probability of finding such a solution to be null. (It is straightforward consequence of the remark after the proof of Theorem 3, in [4]). Therefore, when solving  $(P)$ , unless the initial datum belongs to a set of measure 0, the solution  $u$  to  $(P)$  cannot be expressed with strictly less than  $N - rk(u) + 1$  elements of  $\mathcal{B}$ .

## 4 The total variation

This section is mostly independent of the preceding one. We focus on an approximation of the total variation (TV) defined, for  $u \in \mathbb{R}^{N^2}$  (notice that the ambient space is now of dimension  $N^2$ ), by

$$TV(u) = \sum_{i,j=1}^N f(\nabla u_{i,j}), \quad (9)$$

where  $\nabla u$  stands for a discrete analogue of the gradient of  $u$ , and  $f$  is a norm in  $\mathbb{R}^2$ . We will also assume that the level sets of  $f$  are polyhedron. Such an energy almost satisfies the hypotheses of Theorem 1 and can approximate the standard TV without noticeable change on the results. (Usually, TV is defined with  $f$  equal to the Euclidean norm.) We will see that the meaning of Theorem 1 still makes sense for such an energy.

The first motivation for considering TV comes from its intensive use in image processing (see [11, 12, 13, 14, 2], for few examples on image restoration, and [15, 16, 17], for examples on texture discrimination). With this regard, we would like to mention the well known staircasing artifact (see [13, 12, 18, 19, 20, 21]). This artifact consists in the creation of large homogeneous zones (where  $\nabla u = (0,0)$ ) in the solution to  $(P)$ , when  $E$  equals TV. The link between staircasing and the non differentiability of TV has already been established in [12, 22]. It seems therefore that the notion of rank might relate to staircasing.

Another motivation comes from the fact TV consists in a mixture of two linear analyses (the  $x$  and  $y$  derivatives). Such mixture might be useful in practice. For instance the adaptation of the results we present here to energies which take the form of a sum of  $l^1$  norms in different bases is straightforward. It corresponds to the situation where  $f$  is the  $l^1$  norm in  $\mathbb{R}^2$ .

To make the relation with Theorem 1, the question is whether we can make the link between the rank of an element  $u \in \mathbb{R}^{N^2}$  and some properties of  $u$ . As we will see in the next section, the results are not as instructive as in the previous cases. We are however able to conclude that the rank of  $u$  is large, when  $u$  contains staircasing. Since the constant  $C_K$  in Theorem 1 are a sum over all the  $\bar{u}$  of rank  $K$ , this guaranties that solutions with large staircasing will be over-represented when solving  $(P)$  with  $E = TV$ .

### 4.1 Unit ball of the total variation and polyhedra

Let us go into details. As already said, we consider TV defined by

$$TV(u) = \sum_{i,j=1}^N f(\nabla u_{i,j}),$$

where  $u \in \mathbb{R}^{N^2}$  is an image,  $\nabla u$  stands for a discrete analogue of the gradient of  $u$ , and  $f$  is a norm in  $\mathbb{R}^2$ . In the remaining of the paper, we take for any  $d \in \mathbb{R}^2$

$$f(d) = \sup_{\Psi \in \mathcal{D}} \langle d, \Psi \rangle_2, \quad (10)$$

where  $\langle \cdot, \cdot \rangle_2$  denotes the usual scalar product in  $\mathbb{R}^2$  and  $\mathcal{D}$  is a suitable<sup>1</sup> finite subset of  $\mathbb{R}^2$ . (Notice that, for isotropy reasons, people usually take the Euclidean norm for  $f$ . It is possible to approximate the Euclidean norm very well with functions of the form (10).)

We have, for any  $u \in \mathbb{R}^{N^2}$ ,

$$\begin{aligned} TV(u) &= \sum_{i,j=1}^N \sup_{\Psi \in \mathcal{D}} \langle \nabla u_{i,j}, \Psi \rangle_2 \\ &= \sup_{(\Psi_{i,j})_{1 \leq i,j \leq N} \in \mathcal{D}^{N^2}} \sum_{i,j=1}^N \langle \nabla u_{i,j}, \Psi_{i,j} \rangle_2 \\ &= \sup_{(\Psi_{i,j})_{1 \leq i,j \leq N} \in \mathcal{D}^{N^2}} \sum_{i,j=1}^N u_{i,j} \operatorname{div}(\Psi)_{i,j}. \end{aligned} \tag{11}$$

The  $\operatorname{div}$  operator is simply the adjoint to the  $\nabla$  operator. It goes from  $(\mathbb{R}^2)^{N^2}$  into  $\mathbb{R}^{N^2}$ . We finally obtain

$$TV(u) = \sup_{\varphi \in \mathcal{D}^{TV}} \langle u, \varphi \rangle_{N^2},$$

with

$$\mathcal{D}^{TV} = \left\{ \operatorname{div}(\Psi), \text{ for } \Psi = (\Psi_{i,j})_{1 \leq i,j \leq N} \in \mathcal{D}^{N^2} \right\}.$$

Of course, since  $\mathcal{D}$  is finite,  $\mathcal{D}^{TV}$  is finite and  $\mathcal{L}_{TV}(1)$  is a polyhedron. The fact that TV is not a norm will be addressed once the results of the next section are established.

## 4.2 A link between rank and staircasing

For a given  $u \in \mathbb{R}^{N^2}$ , such that  $TV(u) \neq 0$ , we define the following sets of indexes:

$$I_0 = \{(i, j), f \text{ is differentiable at } \nabla u_{i,j}\},$$

$$I_1 = \{(i, j), f \text{ is not differentiable at } \nabla u_{i,j} \text{ and } \nabla u_{i,j} \neq (0, 0)\},$$

and

$$I_2 = \{(i, j), \nabla u_{i,j} = (0, 0)\}.$$

Of course, we have

$$I_0 \cup I_1 \cup I_2 = \{1, \dots, N\}^2,$$

and

$$I_0 \cap I_1 = I_0 \cap I_2 = I_1 \cap I_2 = \emptyset. \tag{12}$$

---

<sup>1</sup>To prove (13), (14) and (15), we need  $\mathcal{D}$  to contain at least three elements, among which at least two are independent. Also we need  $\langle \Psi, \Psi \rangle_2 = 1$ , for all  $\Psi \in \mathcal{D}$ , and  $\mathcal{D}$  does not contain the same element twice.



**Proposition 4** *Let  $TV$  be defined by (9) and (10). For any  $u \in \mathbb{R}^{N^2}$ , such that  $TV(u) \neq 0$ ,*

$$\#I_2 - \#I_0 \leq rk(u) \leq 2\#I_2 + \#I_1,$$

*where  $\#$  denotes the number of elements of a set.*

Of course, the total variation, as defined by (9) and (10) does not satisfy the hypotheses of Theorem 1, since it is not a norm. However, in order to satisfy the hypotheses of Theorem 1, we can add to  $\mathcal{D}^{TV}$  the elements  $\varepsilon \mathbb{1}$  and  $-\varepsilon \mathbb{1}$ , where  $\varepsilon$  is a small non-negative number and  $\mathbb{1}_{i,j} = 1$ , for all  $(i,j) \in \{1, \dots, N\}^2$ . If we do so,  $rk(u)$  remains unchanged or is increased by one unit. This would not significantly modify the point we want to address.

*Proof.* First, given (10), it is not difficult to see that

$$I_0 = \{(i,j), \exists! \Psi_{i,j}^1 \in \mathcal{D}, f(\nabla u_{i,j}) = \langle \nabla u_{i,j}, \Psi_{i,j}^1 \rangle_2\}, \quad (13)$$

where  $\exists!$  stands for "there exists a unique". Also,

$$I_1 = \{(i,j), \exists! (\Psi_{i,j}^1, \psi_{i,j}^2) \in \mathcal{D}^2, \Psi_{i,j}^1 \neq \Psi_{i,j}^2 \text{ and } f(\nabla u_{i,j}) = \langle \nabla u_{i,j}, \Psi_{i,j}^1 \rangle_2 = \langle \nabla u_{i,j}, \Psi_{i,j}^2 \rangle_2\}, \quad (14)$$

and

$$I_2 = \{(i,j), \forall \Psi \in \mathcal{D}, f(\nabla u_{i,j}) = \langle \nabla u_{i,j}, \Psi \rangle_2\}. \quad (15)$$

For any,  $(i,j) \in I_2$ , we chose two linearly independent elements  $(\Psi_{i,j}^1, \Psi_{i,j}^2) \in \mathcal{D}^2$ .

In order to obtain the upper bound, we consider

$$V \triangleq \{\Psi \in \mathcal{D}^{N^2}, TV(u) = \langle u, \text{div} \Psi \rangle_{N^2}\}.$$

We know that

$$rk(u) = \dim(\text{div}(\text{Span}(V))).$$

It is not difficult to see that

$$V = \{\Psi \in \mathcal{D}^{N^2}, \forall (i,j) \in \{1, \dots, N\}^2, \langle \nabla u_{i,j}, \Psi_{i,j} \rangle = f(\nabla u_{i,j})\},$$

which can be written (using (13), (14) and (15))

$$\begin{aligned} V = \{ \Psi \in \mathcal{D}^{N^2} \quad , \quad & \forall (i,j) \in I_0, \Psi_{i,j} = \Psi_{i,j}^1 \\ & , \quad \forall (i,j) \in I_1, \Psi_{i,j} = \Psi_{i,j}^1 \text{ or } \Psi_{i,j} = \Psi_{i,j}^2 \\ & , \quad \forall (i,j) \in I_2, \Psi_{i,j} \in \mathcal{D} \} \end{aligned}$$

We denote

$$\Psi^1 = (\Psi_{i,j}^1)_{1 \leq i,j \leq N},$$

$\Psi^{2,i,j} \in (\mathbb{R}^2)^{N^2}$ , for  $(i,j) \in I_1 \cup I_2$ , such that

$$\Psi_{i',j'}^{2,i,j} = \begin{cases} (0,0) & , \text{ if } (i',j') \neq (i,j), \\ \Psi_{i,j}^2 - \Psi_{i,j}^1 & , \text{ if } (i',j') = (i,j), \end{cases}$$

and  $\Psi^{1,i,j} \in (\mathbb{R}^2)^{N^2}$ , for  $(i,j) \in I_2$ , such that

$$\Psi_{i',j'}^{1,i,j} = \begin{cases} (0,0) & , \text{ if } (i',j') \neq (i,j), \\ \Psi_{i,j}^1 & , \text{ if } (i',j') = (i,j). \end{cases}$$

We have

$$V \subset \text{Span} \left( \{\Psi^1\} \cup \{\Psi^{2,i,j}\}_{(i,j) \in I_1 \cup I_2} \cup \{\Psi^{1,i,j}\}_{(i,j) \in I_2} \right).$$

So

$$\text{div}(\text{Span}(V)) \subset \text{Span} \left( \{\text{div}(\Psi^1)\} \cup \{\text{div}(\Psi^{2,i,j})\}_{(i,j) \in I_1 \cup I_2} \cup \{\text{div}(\Psi^{1,i,j})\}_{(i,j) \in I_2} \right),$$

which, since  $\text{div}(\Psi^1) \equiv 0$ , implies that

$$rk(u) \leq 2\#I_2 + \#I_1.$$

In order to obtain the lower bound, we denote

$$J(u) \triangleq \{\varphi \in \mathcal{D}^{TV}, TV(u) = \langle u, \varphi \rangle_{N^2}\}$$

and

$$\bar{u} \triangleq \{v \in \mathbb{R}^{N^2}, J(v) = J(u) \text{ and } TV(v) = TV(u)\}.$$

Let us also consider

$$\begin{aligned} W &\triangleq \{\lambda v, \text{ with } \lambda > 0 \text{ and } v \in \bar{u}\}, \\ &= \{v \in \mathbb{R}^{N^2}, J(v) = J(u)\}. \end{aligned}$$

We can easily deduce from the definition of  $J(u)$  that  $\bar{u}$  is an open polyhedron in an affine manifold of dimension  $N^2 - rk(u)$ . (Indeed,  $\bar{u}$  is defined by  $rk(u)$  independent equalities and a consistent system of strict inequalities.) Since  $TV(u) \neq 0$ , it is not difficult to see that  $W$  is included in an affine manifold of dimension  $N^2 - rk(u) + 1$ .

Using again (13), (14), (15), we see that

$$\begin{aligned} W \subset & \{v, \forall (i,j) \in I_0, \langle \nabla v_{i,j}, \Psi_{i,j}^1 \rangle_2 = f(\nabla v_{i,j}) \\ & \forall (i,j) \in I_1, \exists \lambda_{i,j} > 0, \nabla v_{i,j} = \lambda_{i,j} \nabla u_{i,j} \\ & \forall (i,j) \in I_2, \nabla v_{i,j} = (0,0)\} \end{aligned}$$

So,

$$W \subset W' \triangleq \{v, (\nabla v_{i,j})_{1 \leq i,j \leq N} \in W''\},$$

with

$$\begin{aligned} W'' = & \{\varphi \in (\mathbb{R}^2)^{N^2}, \forall (i,j) \in I_2, \varphi_{i,j} = (0,0) \text{ and} \\ & \forall (i,j) \in I_1, \exists \lambda_{i,j} > 0, \varphi_{i,j} = \lambda_{i,j} \nabla u_{i,j}\}. \end{aligned}$$

Notice that  $\dim(W'') = 2N^2 - 2\#I_2 - \#I_1$  and that  $\dim(W'') = \dim(W') + 1$ . So,

$$\begin{aligned} N^2 - rk(u) + 1 &= \dim(W) \\ &\leq \dim(W') + 1 \\ &\leq 2N^2 - 2\#I_2 - \#I_1 + 1. \end{aligned}$$

Using the fact that  $N^2 = \#I_0 + \#I_1 + \#I_2$ , we finally obtain

$$\#I_2 - \#I_0 \leq rk(u),$$

which finishes the proof.  $\square$

Proposition 4 tells us that , when  $u$  is such that  $\#I_2$  is large (strong staircasing)  $rk(u)$  is large too. Indeed, when  $\#I_2$  is large, since

$$\#I_0 + \#I_1 + \#I_2 = N^2,$$

$\#I_0$  is small.

Also, when putting Theorem 1 and Proposition 4 together, we know that if the total variation of a solution  $u$  to a model of the form  $(P)$  is small, it is very likely that  $rk(u)$  is large. It is however not possible to interpret this property in terms of staircasing. The only information we have is that  $\#I_1 + 2\#I_2$  is likely to be large too.

Notice that it is a classical problem in mathematical papers modelling the staircasing artifact in images. The understanding of the staircasing artifact in dimension 2 is not as deep as its counter part for 1 D signals, see [12, 18, 20]. The mostly achieved statements, for images, only establish the stability of homogeneous regions, when the initial datum in the model  $(P)$  already contains staircasing, see [22]. This is of course not in contradiction with the result stated here. Moreover, when adapted to 1 D signals, Proposition 4 makes a clear link between staircasing (in 1 D) and the rank of an element. (The adaptation of the proof of Proposition 4 is straightforward.)

It seems also that the information provided by the rank of an image is not sufficient to establish the high probability of obtaining a strong staircasing in models of the form  $(P)$ . This phenomenon is however observed in many different and independent experiments and can probably be explained. One way to do so might be to improve the analysis provided in [4]. We might find that, for a given rank  $K > 0$ , the contribution to  $C_K$  (see Theorem 1) of elements  $\bar{u}$  for which  $\#I_2$  is large is more important than the contribution of those for which  $\#I_1$  is large.

The proposition however guaranties that solutions with a strong staircasing will be over-represented among solutions to  $(P)$ , when  $E = TV$ , since they contribute to  $C_K$  with  $K$  large.

Finally, if one wants to avoid solutions to  $(P)$  with a large rank (maybe in the expectation of reducing the staircasing artifact), the construction of the

constant  $C_K$  shows that this can be achieved by improving the data fidelity term. However, in this context, Theorem 1 tells us this might be of a limited interest. Of course, we can expect to outperform the results predicted by Theorem 1 when the data fidelity term fails to comply to its hypotheses. Some experiments on the model described in [2] (they have not yet been reported) show the solutions to such models almost do not contain staircasing. As far as we know, the impact of the data fidelity term on staircasing has not yet been exploited by authors working in that field (see [13, 19, 21] and references therein).

### 4.3 Remark

The adaptation of the proof of Proposition 4 to other energies might be simple. For instance, a similar statement holds for an energy of the form

$$E(u) = \sum_{k=1}^{N^2} |\langle u, \varphi_k \rangle_{N^2}|, \quad (16)$$

for  $u \in \mathbb{R}^{N^2}$  and a basis  $(\varphi_k)_{1 \leq k \leq N^2}$  of  $\mathbb{R}^{N^2}$ .

In this case, we denote

$$I_0 = \{k, \langle u, \varphi_k \rangle_{N^2} = 0\}$$

and

$$I_1 = \{k, \langle u, \varphi_k \rangle_{N^2} \neq 0\}.$$

The proposition becomes :

**Proposition 5** *Let  $E$  be defined by (16). For any  $u \in \mathbb{R}^{N^2}$ , such that  $E(u) \neq 0$ ,*

$$rk(u) = \#I_0 + 1$$

The proof is a straightforward adaptation of the proof of Proposition 4.

In this context, Theorem 1 says that, under quite general assumptions on the data distribution law and for a quite general class of data fidelity terms : if the solution  $u$  to  $(P)$ , for  $E$  defined by (16), is such that  $E(u)$  is small,  $u$  is very likely to be sparse, in the basis  $(\varphi_k)_{1 \leq k \leq N^2}$ .

Again, in order to increase the sparsity of the solution in this basis, one should improve the data fidelity term to get better  $C_K$ .

Also, this result suggests it might be possible to extend some theoretical results which are usually stated for an energy of the form (16) to all the energies whose level sets are polyhedral. In such cases, the hypotheses on  $\#I_0$  should be replaced by analogue hypotheses on the rank of  $u$ . (It is, for instance certain that results similar to those presented in [23] can benefit from this generalization.)

Finally, Proposition 5 is only a small improvement on Corollary 1, when  $\mathcal{B} = \tilde{\mathcal{B}} \cup \{-\varphi, \varphi \in \tilde{\mathcal{B}}\}$ , where  $\tilde{\mathcal{B}} = (\tilde{\varphi}_k)_{1 \leq k \leq N^2}$  is the basis such that for all  $u \in \mathbb{R}^{N^2}$ ,

$$u = \sum_{k=1}^{N^2} \langle u, \varphi_k \rangle_{N^2} \tilde{\varphi}_k.$$

## Acknowledgement

I would like to thank Alain Trouvé (CMLA, ENS Cachan, France) for his stimulating remarks and all the help he gave me.

## References

- [1] A. Chambolle, R.A. De Vore, N. Lee, and B.J. Lucier, “Nonlinear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage,” *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 319–355, 1998, Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.
- [2] F. Malgouyres, “Minimizing the total variation under a general convex constraint for image restoration,” *IEEE, trans. on Image Processing*, vol. 11, no. 12, pp. 1450–1456, Dec. 2002.
- [3] F. Malgouyres, “Image compression through a projection onto a polyhedral set,” Tech. Rep. 2004-22, Université Paris 13, August 2004, available at : <http://www.math.univ-paris13.fr/~malgouy>.
- [4] F. Malgouyres, “Projecting onto a polytope simplifies data distributions,” Paper in preparation.
- [5] S. Chen, D. Donoho, and M Saunders, “Atomic decomposition by basis pursuit,” in *SPIE, international Conference on Wavelets*, San Diego, July 1995.
- [6] J-L. Starck, M. Elad, and D.L. Donoho, “Image decomposition via the combination of sparse representations and a variational approach,” *IEEE, trans. on Image Processing*, vol. 14, no. 10, pp. 1570–1582, October 2005.
- [7] M. Brown and N.P. Costen, “Exploratory basis pursuit classification,” *Pattern Recognition Letters*, vol. 26, pp. 1907–1915, 2005.
- [8] D. Donoho, “Neighborly polytopes and sparse solution of underdetermined linear equations,” Tech. Rep. 2005-04, Stanford, Dept of statistics, January 2005.
- [9] D. Donoho and J. Tanner, “Sparse nonnegative solution of underdetermined linear equations by linear programming,” Tech. Rep. 2005-06, Stanford University, April 2005.
- [10] R.T. Rockafellar, *Convex analysis*, Princeton University Press, 1970.
- [11] L. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D*, vol. 60, pp. 259–268, 1992.
- [12] M. Nikolova, “Local strong homogeneity of a regularized estimator,” *SIAM, Journal of Applied Mathematics*, vol. 61, no. 2, pp. 633–658, 2000.

- [13] A. Chambolle and P.L. Lions, “Image recovery via total variation minimization and related problems,” *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.
- [14] F. Malgouyres and F. Guichard, “Edge direction preserving image zooming: a mathematical and numerical analysis,” *SIAM, J. Num. Anal.*, vol. 39, no. 1, pp. 1–37, 2001, A preliminary version is available at: <http://www.math.univ-paris13.fr/~malgouy>.
- [15] Y. Meyer, *Oscillating patterns in image processing and in some nonlinear evolution equation*, AMS, Boston, MA, USA, 2001, The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures.
- [16] S. Osher, A. Solé, and L. Vese, “Image decomposition and restoration using total variation minimization and the  $h^{-1}$  norm,” Tech. Rep. CAM report 02-57, University of California Los Angeles, October 2002, to appear in SIAM MMS.
- [17] J.F. Aujol, G. Aubert, L. Blanc-Ferraud, and A. Chambolle, “Image decomposition into a bounded variation component and an oscillating component,” *Journal of Mathematical Imaging and Vision*, vol. 22, no. 1, pp. 71–88, Jan. 2005.
- [18] W. Ring, “Structural properties of solutions of total variation regularization problems,” Tech. Rep., University of Graz, Austria, 1999, Available at <http://www.kfunigraz.ac.at/imawww/ring/>.
- [19] S. Levine, Y. Chen, and J. Stanich, “Image restoration via nonstandard diffusion,” Tech. Rep. 04-01, Dept mathematics and computer science, Duquesne University, 2004.
- [20] D. Strong and T. Chan, “Edge-preserving and scale-dependent properties of total variation regularization,” *Inverse Problem*, vol. 19, no. 6, pp. 165–187, 2003.
- [21] T. Chan, S. Esedoglu, and F. Park, “A fourth order dual method for staircase reduction in texture extraction and image restoration models,” Tech. Rep., UCLA, 2005.
- [22] M. Nikolova, “Weakly constrained minimization : Application to the estimation of images and signals involving constant regions,” *Journal of Mathematical Imaging and Vision*, vol. 21, pp. 155–175, 2004.
- [23] E. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles : Exact signal reconstruction from highly incomplete frequency information,” Tech. Rep., Caltech, 2004.